ChinaXiv合作期刊

第46卷 第8期
2023年8月

干旱区地理
ARID LAND GEOGRAPHY

Vol. 46 No. 8
Aug. 2023

基于PlanetScope影像的典型绿洲土壤盐渍化数字制图

李 科[1,2]， 丁建丽[1,2]， 韩礼敬[1,2]， 葛翔宇[1,2]，
顾永昇[1,2]， 周 倩[1,2]， 吕阳霞[3]

（1. 新疆大学地理与遥感科学学院/智慧城市与环境建模自治区普通高校重点实验室，新疆 乌鲁木齐 830046；
2. 新疆大学绿洲生态教育部重点实验室，新疆 乌鲁木齐 830046；
3. 新疆阿克苏地区渭干河流域管理局，新疆 阿克苏 842000）

摘 要：干旱半干旱地区急需高分辨率的土壤盐度图用于显示盐度空间分布的细微变化，指导盐渍化区域和潜在盐渍化区域制定土地资源管理政策和水资源管理政策，防止土壤进一步退化，保障农业经济可持续发展和粮食安全生产。基于PlanetScope影像，提取植被光谱指数和土壤盐度指数，共计21个变量，将其输入装袋回归（Bootstrap aggregating，Bagging）算法中，构建了土壤盐度预测模型Model-Ⅰ；使用最相关最小冗余（Max-relevance and min-redundancy，mRMR）方法筛选特征变量，将其输入Bagging中，构建了土壤盐度预测模型Model-Ⅱ，使用野外采样数据来辅助建模并进行验证。通过模型评价指标对Model-Ⅰ和Model-Ⅱ进行评估。结果表明：Model-Ⅱ的预测性能优于Model-Ⅰ（验证集决定系数为0.66，均方根误差为18.00 dS·m^{-1}，四分位数的相对预测误差为3.21），mRMR有效降低了多维特征冗余问题。PlanetScope影像结合mRMR方法成功绘制了高分辨率土壤盐度图，提供了更详细的土壤盐度空间分布信息，研究结果对利用PlanetScope数据监测土壤盐渍化信息起推动作用。

关 键 词：PlanetScope；mRMR；Bagging；土壤盐渍化；数字制图技术
文章编号：1000－6060(2023)08－1291－12(1291～1302)

chinaXiv:202309.00163v1

土壤盐渍化阻碍了植被对水分的吸收，加速了土壤表面有机质分解，降低了土壤肥力和土壤团聚体稳定性，是土壤退化的主要因素之一[1]。全球范围内有7.7%~11.5%的灌溉土地受到盐渍化威胁，这意味着过量的可溶性盐分累积，对作物生产和农业可持续发展带来严重影响[2]。我国发生盐渍化的地区分布广泛，其中新疆地区的盐渍土类型最多，盐渍土总面积达847.6×10^4 hm^2，占新疆地区现有耕地面积的31.1%[3]。充足的耕地资源是保障粮食生产和安全的基石，但不断加剧的土壤盐渍化每年将1%~2%的耕地转化为盐渍地，给粮食生产带来了严重限制和挑战[4]。土壤盐度是评估盐渍化严重程度的有效指标，但在水平和垂直方向呈现出快速的时空动态变化[5]。因此，通过制作土壤盐度分布图能帮助决策者了解土壤盐度存在的巨大时空差异，为潜在盐渍化区域制定优化的土地资源管理和水资源管理政策，为盐渍化区域制定优化的土壤改良政策，从而应对新疆地区土地持续性退化难题[6]。

20世纪70年代，新疆地区通过第二次全国土地调查获取了土壤盐度数据，使用矢量空间数据结构记录其空间信息，这种数据记录方式难以准确地反映土壤属性的空间异质性，也难以捕获盐度信息的动态变化，无法满足新疆地区对土壤盐渍化管理和预防的需求[7]。研究证明，光谱学知识结合遥感技

收稿日期：2022-09-30； **修订日期：**2022-10-29
基金项目：国家自然科学基金项目（41961059）；新疆维吾尔自治区自然科学基金重点项目（2021D01D06）；新疆维吾尔自治区研究生科研创新项目（XJ2022G002）资助
作者简介：李科（1996-），男，硕士研究生，主要从事干旱区遥感与GIS应用研究. E-mail: dixinlike@stu.xju.edu.com
通讯作者：丁建丽（1974-），男，博士，教授，主要从事干旱区遥感与GIS应用研究. E-mail: watarid@xju.edu.cn

ChinaXiv合作期刊

术可以快速且定量地获取土壤盐度信息,并成功用于绘制高分辨率的土壤盐度图,提供了更全面的土壤盐渍化空间信息[8]。盐渍化土壤相较于非盐渍化土壤,在可见光到短波红外波长范围内具有更高的光谱反射率,随着土壤盐度增加,光谱反射率进一步提高,在近红外波长附近出现具有诊断性的吸收带[9-11]。因此,归一化盐度指数(Normalized difference salinity index,NDSI)和多种盐度指数(Salinity index,SI)被广泛用于评估土壤盐度。植物对土壤盐碱变化的耐受,通过生长状况进行响应,长势不好的植被光合作用减弱,叶绿素含量变少,可见光范围内反射率增强,近红外范围内反射率降低[12]。因此,归一化植被指数(Normalized differential vegetation index,NDVI)和土壤调节植被指数(Soil adjusted vegetation index,SAVI)等多种植被指数(Vegetation indices,VIs)被广泛用于评估土壤盐度的间接指标。Gorji等[13]基于Landsat8 OLI和Sentinel-2A MSI影像,成功提取了与土壤电导率相关的SI,并结合多元线性回归技术成功绘制了乌尔米耶湖流域土壤盐度图。Fernandez-Buces等[14]基于Lansat ETM、机载近红外和实测光谱数据获取了不同植被和裸地的反射率,提取并调整了NDVI,并计算了综合光谱指数(Combined spectral response index,COSRI),以制作出高精度的原墨西哥特斯科科湖区域盐度图。Ding等[15]使用近端遥感装置EM38测量表观电导率数据,基于Landsat TM影像提取了SI、NDSI、亮度指数(Britness index,BI)、NDVI和差值植被指数(Difference vegetation index,DVI),将表观电导率数据和遥感指数信息应用于回归克里金插值法,成功评估了土壤盐度并获得了满意的结果。

20世纪末,机器学习算法被广泛使用,在土壤科学领域,基于机器学习算法绘制了多种数字土壤地图,如Farifteh等[16-17]基于随机森林方法(Random forest,RF)开发了绘制土壤有机碳浓度和储量数字图的框架。Ma等[18-19]基于RF和极端梯度提升等机器学习方法绘制了渭库绿洲土壤盐度图,有效解决了土壤和环境因子间的非线性问题。Wang等[20]在绘制不同区域的土壤盐度图时,比较了5种机器学习算法,发现基于机器学习算法进行土壤盐度预测研究时需要用户定义的参数更少,拥有更高的计算效率,能够处理数值序数和离散预测变量等优点。然而,在构建高效、准确和泛化程度高的土壤盐度预测模型时,特征变量的选择至关重要。Allbed等[19-21]的研究结果表明植被指数和盐分指数等特征变量可以有效用于土壤盐渍化监测研究,但是这些特征变量均是通过波段间的运算得到,造成了特征变量间不同程度的信息冗余。最相关最小冗余(Max-relevance and min-redundancy,mRMR)特征选择算法具有较低的复杂度,与当前常用的特征选择算法相比具有原理简单和对数据要求低的优点,mRMR具有的优势提升了土壤盐度预测模型的预测效率[22]。Peng等[23]使用Pearson相关分析筛选相关性高的特征变量进行土壤盐渍化监测,并没有考虑选择特征间存在的信息冗余。马慧琴等[24]在筛选特征用于构建遥感监测模型时,对比了mRMR与常用相关分析方法,研究结果表明mRMR方法比常用相关分析方法更具优势。

目前,遥感影像的中低空间分辨率仍然是影响土壤盐度图精度的主要因素[25]。当受盐影响区域小于像素大小时,会限制土壤盐度图对土壤盐度进行准确评估的能力,空间分辨率低于5 m的多光谱传感器正成为绘制土壤盐度图不可或缺的装置[26]。Bouaziz等[27]证实了基于MODIS terra数据提取的光谱指数与实测数据能有效绘制土壤盐度图,但较低的空间分辨率是造成二者相关性较差的主要原因。Eldeiry等[28]比较了IKONOS与Landsat影像在农业地区评估土壤盐度的性能,结果表明高空间分辨率能获取更准确的评估结果。自2016年以来,Planet公司运营的地球成像小型卫星星座PlanetScope,能够以较高的空间分辨率(<5 m)和时间分辨率(1~2 d)获取影像数据。在2019年4月PlanetScope携带了第三代传感器PSB.SD能获取可见光近红外范围内的8个波段[29]。PlanetScope获取的影像与Landsat8 OLI[30]和Sentinel-2A MSI[31]相比,具有更高的空间分辨率和时间分辨率,而且国内利用PlanetScope小型星座卫星影像进行土壤盐度制图的研究并未广泛展开。促进高空间和高时间分辨率的遥感影像用于土壤盐度监测研究,能获取更翔实的空间信息、动态变化的盐度信息和更精确的评估结果。

本文以渭干河-库车河流域三角洲绿洲(简称渭库绿洲)为研究区,基于PlanetScope卫星的原始波段数据和波段运算所得的光谱指数,共21个特征变量。采用mRMR方法筛选重要特征作为自变量

chinaXiv:202309.00163v1

ChinaXiv合作期刊

chinaXiv:202309.00163v1

输入Bagging回归算法，建立了土壤盐度预测模型，成功绘制了研究区高分辨率土壤盐度分布图。这为新疆地区制定管理改善土壤盐碱策略，维持绿洲可持续发展和保障粮食安全提供重要参考。

1 材料与方法

1.1 研究区概况

渭库绿洲，属于南疆绿洲区。地势北高南低、地貌复杂多样，气候干燥、蒸发强烈、降水稀少，属于典型的温带大陆性干旱气候[15]。复杂的地形地貌、极端的气候条件、农业上排水系统不健全、灌溉技术粗放等问题[32]，导致研究区存在不同程度的盐渍化。绿洲外围植被稀疏，生态系统脆弱，主要包括芦苇（*Phragmites australis*）、柽柳（*Tamarix chinensis*）、碱蓬（*Suaeda glauca*）和骆驼刺（*Alhagi sparsifolia*）等植物。实际研究区位于库车市绿洲外围农田荒漠交错带，地理位置介于83°23′~83°29′E、41°23′~41°27′N之间（图1）。

1.2 样品采集与处理

实验团队于2021年6月23—24日在研究区进行了实地调查，结合以往的实地调查成果和采样经验，在农田荒漠交错带选择了84个具有代表性的采样点。用便携式GPS记录采样点的地理位置，以采样点为中心，在距离采样点1 m的范围内，以0~10 cm的深度使用土壤取样器均匀的采集5个样品，然后将5个样品混合在一起，取混合后的样品约500 g，装入有标签的防水密封袋。在实验室里，所有的土壤样品均经过风干、研磨和过筛（2.0 mm）等一系列处理。取20 g处理过的土样和100 mL纯水充分振荡，静置24 h。在25 ℃的温度下，用多参数测量仪（WTWinoLab® Multi3420 set B，WTW GmbH，Germany）测量静置后的浸出液，得到电导率（EC）值。由于土壤中的盐分溶解在浸出液中，浸出液的含盐量与25 ℃下电极的导电能力呈正相关关系，通过测量土壤浸出液的EC值，可以作为评估土壤盐度的参考指标[33]。

1.3 遥感影像预处理与光谱指数提取

PlanetScope是通过多组发射的单独卫星组成的卫星星座，目前约有130颗CubeSat 3U（10×10×30 cm）卫星[34]。目前已经搭载了第3代传感器PSB.SD，幅宽为32.5×19.6 km，能采集8波段的多光谱影像，空间分辨率为3 m。这些卫星大多位于太阳同

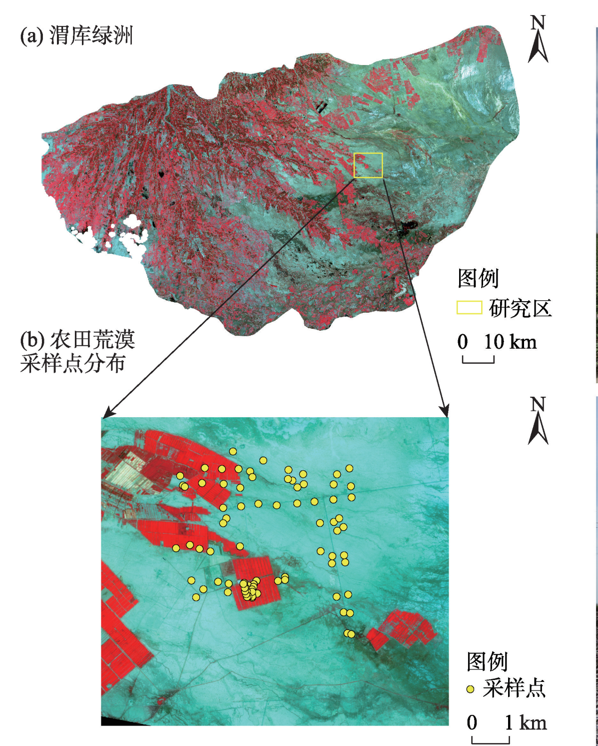






图1　研究区概况图

Fig. 1　Overview map of the study area

ChinaXiv合作期刊

步轨道，所有卫星组成的卫星星座PlanetScope拥有每天对地球成像的能力。本研究使用的PlanetScope波段参数见表1，使用蓝波段(B2)、绿波段Ⅱ(B4)、红波段(B6)、红边波段(B7)和近红外波段(B8)进行了后续研究。本文使用的是PlanetScope Level-3B数据，过境时间为2021年7月24日，Level-3B产品使用6S辐射传输模型和MODIS数据作为辅助对PlatnetScope获取的遥感影像进行大气校正，并经过了传感器校正和正射校正。

表1 PlanetScope影像光谱波段参数

Tab. 1 Spectral band parameters of PlanetScope images

波段	波段名	波长/nm
蓝波段	B2	465~515
绿波段Ⅰ	B3	513~549
绿波段II	B4	547~583
黄波段	B5	600~620
红波段	B6	650~680
红边波段	B7	697~713
近红外波段	B8	845~885

将光谱指数作为预测土壤特性模型的输入变量时，能有效降低光谱反射率误差[35]。通过波段运算得到的盐度指数与植被指数被证明能准确反演土壤盐分[36]。本研究使用PlanetScope原始波段数据计算光谱指数，计算公式如表2所示。

1.4 mRMR特征变量选择算法

mRMR是一种过滤型特征选择算法，其算法原理简单且对数据的要求低，在土壤盐度监测领域还没有相关的研究探讨该算法的应用。因此，本研究结合mRMR方法构建了土壤盐度预测模型，为土壤盐度监测领域提供了参考。mRMR考虑了特征之间相似性造成的数据冗余问题，通过计算特征与离散化电导率数据类别之间的互信息，特征与特征之间的互信息，来度量它们之间的相关性。然后，保留与离散化电导率数据类别之间互信息值较高的特征，同时去除特征之间互信息值较高导致数据冗余的特征[44]。任意2个特征x和y的互信息计算公式为：

$$I(x,y)=\iint P(x,y)\log\frac{P(x,y)}{P(x)P(y)}\mathrm{d}x\mathrm{d}y \tag{1}$$

式中：$I(x,y)$为特征x和y的互信息；$P(x,y)$为特征x和y的联合概率密度函数；$P(x)$为特征x的边缘概率密度函数；$P(y)$为特征y的边缘概率密度函数。特征与离散化电导率数据类别之间的互信息计算公式为：

表2 基于PlanetScope影像光谱指数计算

Tab. 2 Spectral indices calculation based on PlanetScope images

光谱指数	名称	公式	参考
土壤调节植被指数	SAVI	SAVI = 1.5×(NIR − Red)/(NIR + Red + 0.5)	[37]
比值植被指数	RVI	RVI = NIR/Red	[38]
增强植被指数	EVI	EVI = 2.5×(NIR − Red)/(NIR + 6×Red − 7.5×Blue + 1)	[39]
归一化差值植被指数	NDVI	NDVI =(NIR − Red)/(NIR + Red)	[40]
归一化差异盐度指数	NDSI	NDSI =(Red − NIR)/(Red + NIR)	[41]
亮度指数	BI	$BI=\sqrt{Red^2+NIR^2}$	[41]
盐度指数	SI	$SI=\sqrt{Blue\times Red}$	[41]
盐度指数	SI1	$SI1=\sqrt{Green\times Red}$	[42]
盐度指数	SI2	$SI2=\sqrt{Green^2+Red^2+NIR^2}$	[42]
盐度指数	SI3	$SI3=\sqrt{Green^2+Red^2}$	[42]
盐度指数	S1	S1 = Blue/Red	[43]
盐度指数	S2	S2 =(Blue − Red)/(Blue + Red)	[43]
盐度指数	S3	S3 =(Green × Red)/Blue	[43]
盐度指数	S4	$S4=\sqrt{Blue\times Red}$	[43]
盐度指数	S5	S5 =(Blue × Red)/Green	[43]
盐度指数	S6	S6 =(Red × NIR)/Green	[43]

注：Red为红波段；Blue为蓝波段；Green为绿波段；NIR为近红外波段。

chinaXiv:202309.00163v1

ChinaXiv合作期刊

chinaXiv:202309.00163v1

$$\begin{cases} D(S,p)=\dfrac{1}{|S|}\sum_{x_i\in S} I(x_i,p) \\ \max D(S,p) \end{cases} \tag{2}$$

式中：$D(S,p)$为特征与类别之间的平均互信息；S为变量集合；$I(x_i,p)$为特征i与类别p之间的互信息；x_i为特征i；p为离散化电导率数据类别。特征与特征之间的互信息计算公式为：

$$\begin{cases} R(S)=\dfrac{1}{|S|^2}\sum_{x_i,x_j\in S} I(x_i,x_j) \\ \min R(S) \end{cases} \tag{3}$$

式中：$R(S)$为特征与特征间的平均互信息；$I(x_i,x_j)$为特征i与特征j之间的互信息；x_i为特征i；x_j为特征j。结合公式(2)与公式(3)得到最大相关最小冗余的判断条件商标准：

$$\begin{cases} \varphi(D,R)=D/R \\ \max \varphi(D,R) \end{cases} \tag{4}$$

式中：$\varphi(D,R)$为互信息商标准；D为特征与类别之间的互信息；R为特征与特征间的互信息。mRMR算法通过Matlab 2019b运行fscmrmr函数实现，fscmrmr函数自动将电导率数据离散化处理。

1.5 建模策略及评价

1.5.1 Bagging回归算法 装袋算法(Bootstrap aggregating, Bagging)是Breiman[45]提出的一种并行集成学习方法，将多个单一的预测器聚合在一起，对结果进行平均处理，进行分类或者回归预测相关的研究。Bagging在训练集中进行Bootstrap采样，即从m个样本中进行随机的、有放回的采样，得到n个与原始训练集样本个数相等的采样集。单一预测器基于这些采样集进行独立的训练，在进行回归分析时，将n个预测器的均值作为综合预测结果，过程如图2所示。Bagging回归算法有效降低了训练集的方差，避免预测结果出现过拟合。通过Python 3.9导入机器学习库sklearn构建Bagging回归模型。

1.5.2 模型构建及评价 并非所有特征都能提升Bagging回归算法预测性能，不相关特征和冗余特征往往会降低模型预测的精度和效率。为了降低负面影响，探索PlanetScope多光谱数据预测土壤盐分的潜力，所有特征构成集合Ⅰ，mRMR算法选择前30%的特征构成集合Ⅱ，基于集合Ⅰ和集合Ⅱ，开发了两种模型Model-Ⅰ和Model-Ⅱ(表3)。将采集的84个土壤样品进行随机划分(在Python 3.9中test_size=0.3用于划分数据集)。所有样品的70%，约58个土样被划分为训练集，用于训练模型；所有样品的30%，约26个土样被划分为验证集，用于模型的验证。为了确保模型的稳健性，提高模型的适用性，降低随机划分样本对模型预测性能的影响，本研究随机划分土壤样本3次，获得3组训练集和验证集(A、B和C)。为了确保模型的可比性，每次划分训练集和验证集时，对于土壤样本与集合Ⅰ和

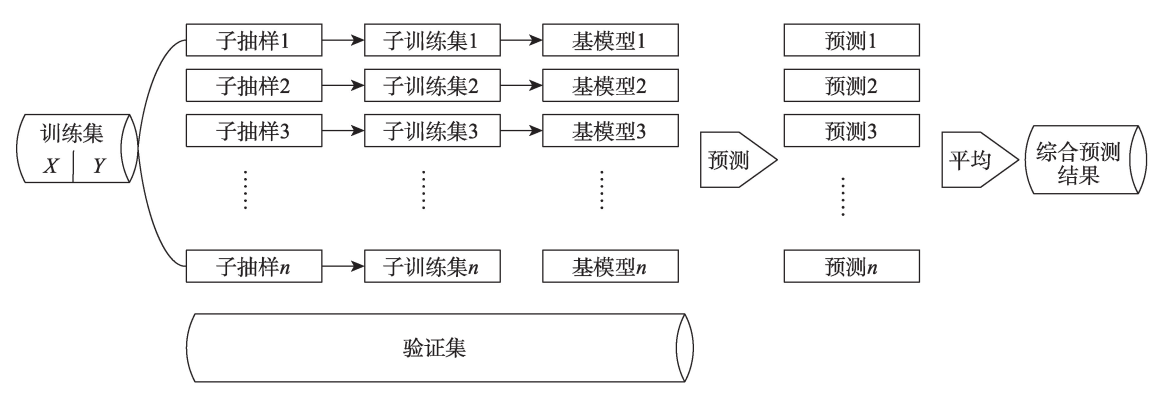


图2　Bagging回归算法预测过程

Fig. 2　Prediction process of Bagging regression algorithm

表3　PlanetScope数据建模策略细节

Tab. 3　Details of the PlanetScope modeling strategies

模型	模型方法	特征选择	特征变量类型
Model-Ⅰ	Bagging	-	波段反射率数据(B2、B4、B6、B7、B8)+光谱指数(SAVI、RVI、EVI、NDVI、NDSI、BI、SI、SI1、SI2、SI3、S1、S2、S3、S4、S5、S6)
Model-Ⅱ	Bagging	mRMR	波段反射率数据(B4)+光谱指数(EVI、S6、S1、SI2、NDSI、BI)

ChinaXiv合作期刊

集合Ⅱ构成的数据集，划分模式保持一致。

为了评估Bagging回归模型预测EC的能力，使用决定系数(R^2)、均方根误差(RMSE)和四分位数的相对预测误差(RPIQ)3个指标进行衡量。模型预测的准确性与R^2、RPIQ成正比，与RMSE成反比，即R^2与RPIQ越大，RMSE越小模型预测性能越好，计算公式如下：

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{p}_i - p_i)^2}{\sum_{i=1}^{n}(p_i - \bar{p})^2} \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{p}_i - p_i)^2} \tag{6}$$

$$\mathrm{RPIQ} = \frac{\mathrm{IQR}}{\mathrm{RMSE}_P} \tag{7}$$

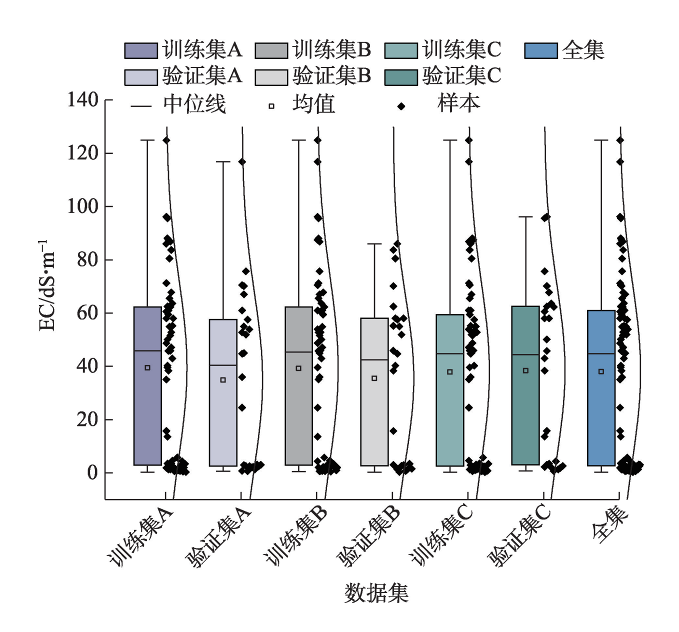
$$\mathrm{IQR} = Q3 - Q1 \tag{8}$$

式中：$\hat{p}_i$ 为预测值；p_i为实际测量值；$\bar{p}$ 为测量值的平均值；n 为采样点数量；IQR为验证集中测量值的四分位间距；RMSE_P为验证集的均方根误差；$Q3$ 为上四分位值；$Q1$ 为下四分位值。

2 结果与分析

2.1 土壤电导率描述性统计

EC作为土壤盐分的参考，EC的描述性统计结果如图3和表4所示。整个EC的范围为0.35~124.90 dS·m^{-1}，平均值为38.04 dS·m^{-1}，标准差为33.13 dS·m^{-1}，变异系数为0.87，为中等变异。EC描述性统计结果显示，土壤样本随机划分3次得到A、B和C数据集，它们的训练集和验证集分布与整个数据集的分布相似。因此，训练集和验证集的样本可以代替整个数据集的样本进行建模和验证。



注：EC为电导率。下同。

图3 电导率统计图

Fig. 3 Statistical diagram of electrical conductivity

表4 数据集A、B和C与全集EC描述性统计

Tab. 4 Descriptive statistics of datasets A, B, and C compared to the entire dataset EC

数据集	平均值/dS·m^{-1}	标准差/dS·m^{-1}	变异系数
训练集A	39.47	33.55	0.85
验证集A	34.86	32.60	0.93
训练集B	39.19	34.45	0.88
验证集B	35.51	30.48	0.86
训练集C	37.88	33.82	0.89
验证集C	38.42	32.21	0.84
全集	38.04	33.13	0.87

2.2 特征变量优选

集合Ⅰ由5个波段反射率数据B2、B4、B6、B7和B8，4个植被光谱指数RVI、EVI、SAVI和NDVI，12个土壤盐度光谱指数S1、S2、S3、S4、S5、S6、SI、SI1、SI2、SI3、BI和NDSI组成。从集合Ⅰ中选择最能反映土壤盐渍化发生发展状况的特征变量构建土壤盐渍化预测模型，能有效提升模型效率和预测精度。目前mRMR特征选择方法还未广泛用于土壤盐渍化监测领域，本研究使用Matlab 2019b统计和机器学习工具箱执行fscmrmr函数，通过mRMR算法计算互信息，并依据互信息商标准保留最相关特征，过滤冗余特征。经过互信息商标准选择前30%特征作为集合Ⅱ，包括1个波段反射率数据B4，1个植被光谱指数EVI，4个土壤盐度光谱指数S6、S1、SI2和NDSI，图4为波段反射率与光谱指数特征重要性。

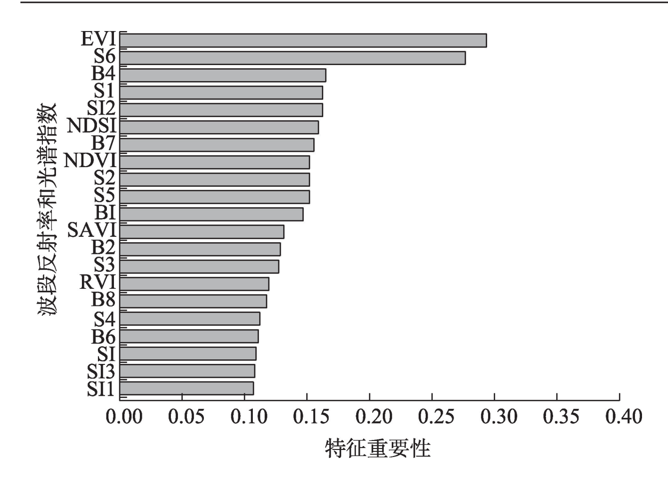
2.3 预测模型和性能评估

为了提高模型的预测精度和泛化程度，使用基于步长为1的网格搜索，对模型影响较大的基础学习器数量进行了1000次连续迭代调整，根据模型R^2值确定模型参数最优值。在进行参数优化时，其余参数为默认值。图5显示了模型R^2与模型参数在连续迭代过程中的变化关系，当R^2最大时确定基础学习器数量数值，重复此过程直到确定参数最优值，Model-Ⅰ模型基础学习器数量为4，Model-Ⅱ模型基础学习器数量为9。

chinaXiv:202309.00163v1

ChinaXiv合作期刊

chinaXiv:202309.00163v1

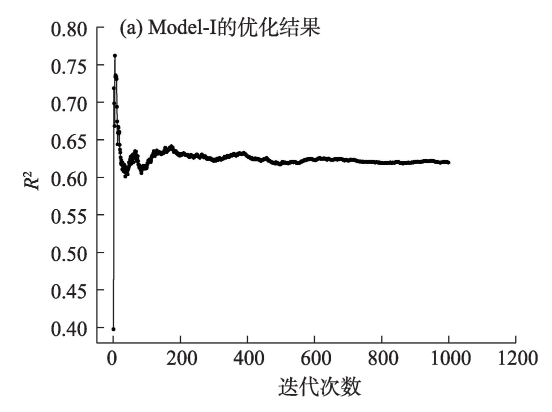


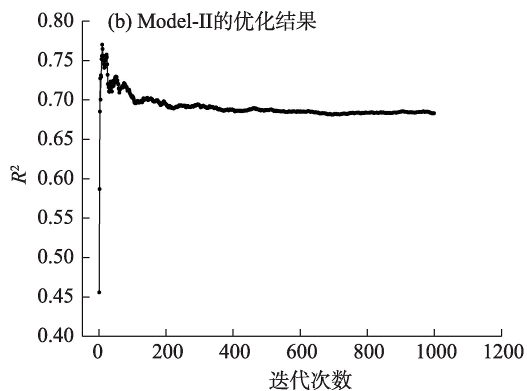
注：SAVI为土壤调节植被指数；RVI为比值植被指数；EVI为增强植被指数；NDVI为归一化差值植被指数；NDSI为归一化差异盐度指数；BI为亮度指数；SI、SI1、SI2、SI3、S1、S2、S3、S4、S5和S6为盐度指数。

图4　PlanetScope波段反射率和光谱指数的特征重要性

Fig. 4　Importance of PlanetScope band reflectance and spectral indices

基于优化后的参数，将集合Ⅰ和集合Ⅱ随机划分的训练集作为自变量输入Bagging模型，构建EC预测模型Model-Ⅰ和Model-Ⅱ，利用随机划分的验证数据集对模型的预测性能进行验证。模型Model-Ⅱ的R^2与RPIQ均高于Model-Ⅰ，RMSE均低于Model-Ⅰ，这说明mRMR算法可以选择更重要的特征代替全部特征，降低数据的复杂性和冗余性，建立更准确和更高效的模型（表5）。在随机划分的三组数据集A、B和C中，Model-Ⅱ与Model-Ⅰ相比，R^2平均提高了27%，RPIQ平均提高了20%，RMSE平均降低了16%。mRMR在选择特征数量较少的条件下，能保留最相关最小冗余特征，能最大化特征包含的有用信息。在集合Ⅱ中光谱指数EVI和S6重要性最高，表明EVI对土壤盐渍化引起的植被状况与光谱反射率变化敏感，S6对盐渍土表面状况与光谱反射率变化敏感。光谱指数EVI与S6在土壤盐渍化监测中具有巨大潜力。





注：R^2为决定系数。下同。

图5　模型参数调整

Fig. 5　Model parameters adjustment

表5　模型预测性能比较

Tab. 5　Comparison of model prediction performance

数据集	模型	训练集		验证集		RPIQ
		R^2	RMSE/dS·m^{-1}	R^2	RMSE/dS·m^{-1}	
A	Model-Ⅰ	0.79	15.26	0.67	18.41	3.34
	Model-Ⅱ	0.90	10.27	0.76	15.78	3.93
B	Model-Ⅰ	0.79	15.82	0.41	22.95	2.27
	Model-Ⅱ	0.87	12.48	0.66	17.45	3.00
C	Model-Ⅰ	0.76	16.43	0.49	22.58	2.42
	Model-Ⅱ	0.87	11.93	0.57	20.75	2.71
平均	Model-Ⅰ	0.78	15.84	0.52	21.32	2.68
	Model-Ⅱ	0.88	11.56	0.66	18.00	3.21

注：R^2为决定系数；RMSE为均方根误差；RPIQ为四分位数的相对预测误差。

ChinaXiv合作期刊

2.4 土壤盐度空间分布

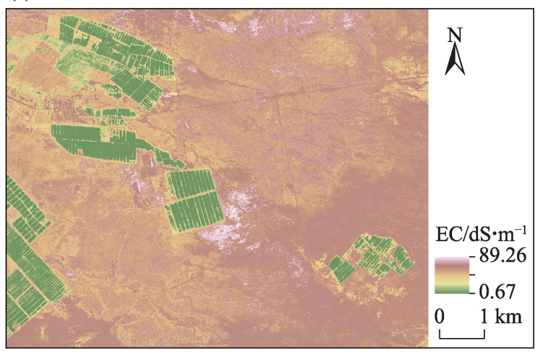
基于模型Model-Ⅰ和Model-Ⅱ,使用随机划分的数据集A、B和C绘制了3张高分辨率的土壤EC分布图,对模型Model-Ⅰ和Model-Ⅱ的3张预测图进行平均,得到图6a和图6b。图6a和图6b产生了相似的土壤EC空间分布,Model-Ⅱ模型产生了最理想的结果,预测的土壤EC范围(0.70~100.83 dS·m^{-1})更符合实际测量的土壤EC值。图6b相较图6a,更准确的显示了农田区域与荒漠区域土壤EC分布的细微变化,更符合该区域土壤盐度分布。受地形影响,土壤盐度高值主要分布在荒漠区域;受地形和灌溉的影响,土壤盐度低值主要分布在农田区域。

3 讨 论

PlanetScope数据结合mRMR特征选择方法首次用于绘制高空间分辨率的土壤EC图,土壤盐渍化监测领域缺乏类似的研究作为参考,因此借鉴了马慧琴等[24]和Jing等[44]的研究,通过比较模型Model-Ⅰ和Model-Ⅱ平均预测性能(表5),得出与其相似的研究结果,mRMR方法能优化建模特征并改善模型预测性能。基于PlanetScope数据建立的模型Model-Ⅰ和Model-Ⅱ获取的预测值如图7所示,Model-Ⅱ获取的预测值与实测值更接近1:1拟合线,表明Model-Ⅱ能获取更准确的预测结果。基于PlanetScope影像低于5 m的空间分辨率,能避免如Scudiero等[46]使用Landsat7多光谱影像进行土壤盐度评估时,结果难以表征土壤盐度的空间变异性等情况。也能避免使用MODIS系列、Landsat系列或Sentinel系列数据进行土壤盐度监测时,预测结果难以具备指导意义等情况[47]。基于此,使用PlanetScope数据结合mRMR特征选择方法建立的模型Model-Ⅱ对研究区土壤EC进行预测并绘制成图,避免了由于空间分辨率不足对土壤盐度监测和绘制带来的限制。

本研究将光谱指数作为自变量输入Bagging模型,改善模型的预测性能,提高模型定量描述土壤盐度的能力,丁建丽等[48]与曹雷等[49]的研究结果证实这是可行的。并非所有光谱指数都能用于优化模型,其中不相关和冗余的特征变量反而会起反作用。本文通过mRMR方法对所有特征变量进行了重要性排序,将排序后的特征变量作为自变量输入模型Model-Ⅱ中。当选择的特征变量数小于30%时,模型的预测精度较低;当选择的特征变量数超过30%时,模型的预测性能略低基本保持稳定。

重要性排序中EVI重要性最高,远远高于其他波段与光谱指数(图4),对土壤盐渍化具备更多的解释能力。这可能是因为植物叶面积和冠层结构的发育受到土壤盐渍化抑制更明显,而EVI对于植物冠层结构变化更敏感,故EVI在建模时更重要。Rao等[50]的研究表明,盐渍土具备更高的光谱反射率,且在可见光和近红外波长范围,盐渍化程度与光谱反射率成正比,在盐渍化程度较高的农田荒漠接壤处与荒漠区域可以直接利用PlanetScope原始波段反射率对土壤盐度进行建模分析。mRMR选择的特征集合Ⅱ中仅包含原始波段B4,故B4相较于其他原始波段,对于土壤盐渍化更敏感,更适合用于描述土壤盐度信息。在本研究中除植被覆盖度较高的农田以外,均能使用对土壤盐度敏感的盐



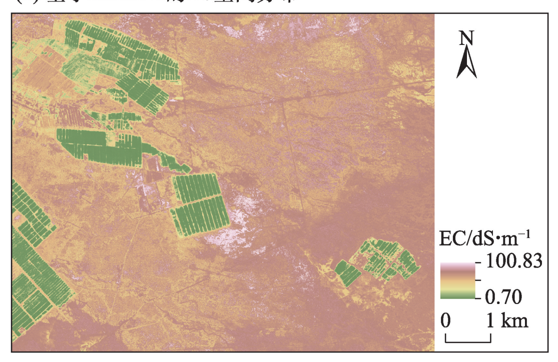
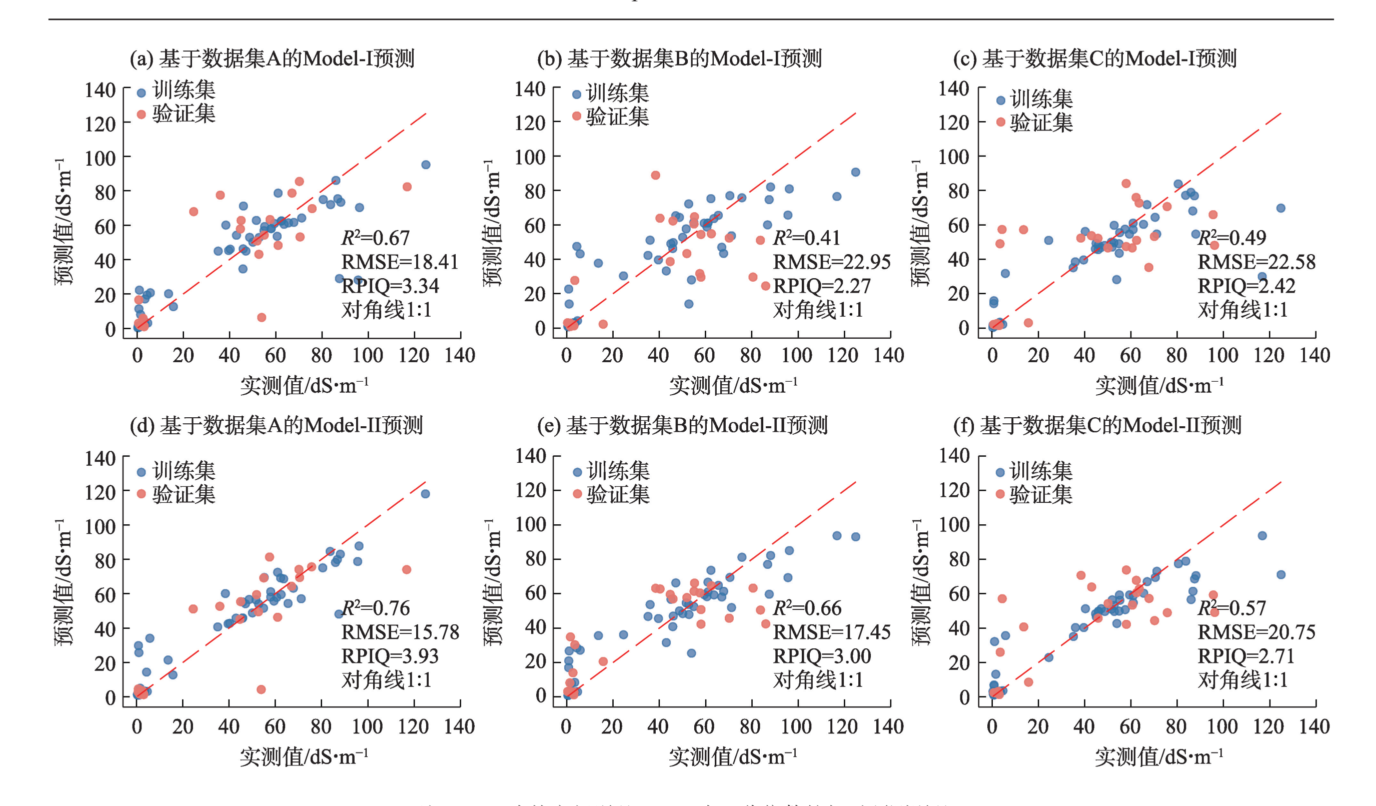


图6 土壤电导率空间分布

Fig. 6 Spatial distributions of soil electrical conductivity

chinaXiv:202309.00163v1

ChinaXiv合作期刊



注：RMSE为均方根误差；RPIQ为四分位数的相对预测误差。

图7　模型Model-Ⅰ和Model-Ⅱ预测性能比较

Fig. 7　Comparison of prediction performance of model Model-Ⅰ and Model-Ⅱ

度指数定量描述土壤盐渍化情况，mRMR选择的前30%特征变量中，S6、B4、S1和NDSI均为盐度指数，故盐度指数对于建立土壤盐度预测模型必不可少。冯娟等[51]、Metternicht等[52]和丁建丽等[53]的研究结果证明了这些观点是合理的。

土壤样本的随机划分和模型参数的设定使模型的误差出现累积。随机划分的训练集和验证集与整个数据集的分布相似，可以代替整个数据集的样本进行建模和验证。为避免样本随机划分为模型带来的不确定性，将随机划分3次得到的数据集A、B和C用于训练模型和验证模型，最终对评价指标取均值来评估模型，增加了模型预测的稳健性，Jing等[44]的研究结果证明这是有效的。虽然多次随机划分样本能有效降低模型预测的不确定性，但是本研究并没有确定最有效的随机划分次数。在设定模型参数时，为了降低模型参数对预测性能的影响，基于步长的网格搜索，对模型影响较大的基础学习器数量进行了连续迭代调整，结果如图5所示。基于步长的网格搜索是否是参数优化的最佳方式，这也存在不确定性。除上述两点，人类活动也增加了模型预测的不确定性，图6a和图6b左上角区域显示，通过Model-Ⅰ和Model-Ⅱ预测的农田区域土壤盐度被高估，这与实际情况不符。通过野外调查验证，该区域为了达到节水增产的目的，推广了膜下滴灌技术。覆膜区域光谱反射率不同于裸土区域，整体偏高，由于多光谱影像的光谱分辨率不足，使得盐碱地与覆膜地难以进行区分，在未来的研究中将尝试利用高光谱技术来消除农田区域覆膜给土壤盐度监测带来的影响。此外，本研究仅考虑了Bagging回归模型，马国林等[54]的研究结果表明，仅通过一个模型难以取得最佳预测效果。因此，在未来的研究中将构建多种模型对土壤盐度进行预测，并对比模型性能选择最适合本研究数据集的模型。

4　结 论

本研究基于PlanetScope影像原始波段数据、植被光谱指数和土壤盐度指数共21个特征变量，使用mRMR方法筛选前30%特征变量并输入Bagging回归模型，建立Model-Ⅰ和Model-Ⅱ土壤盐度预测模型，成功绘制了研究区3 m分辨率的EC空间分布图，为准确监测随时空动态快速变化的土壤盐度信

chinaXiv:202309.00163v1

ChinaXiv合作期刊

息提供参考。得出以下结论：

（1）将PlanetScope影像原始波段数据、植被光谱指数和土壤盐度指数共21个特征变量作为自变量建立的Model-Ⅰ模型，基本可以实现对EC的预测（验证集平均R^2=0.52，平均RMSE=21.32 dS·m^{-1}，平均RPIQ=2.68）。

（2）将mRMR筛选后的特征变量作为自变量建立的Model-Ⅱ模型，降低了预测EC的特征维度，提高了模型的预测效率和准确性（验证集平均R^2=0.66，平均RMSE=18.00 dS·m^{-1}，平均RPIQ=3.21）。Model-Ⅱ与Model-Ⅰ相比，EVI、S6、B4、S1、SI2和NDSI对农田荒漠交错带土壤盐渍化监测具有重要作用。

参考文献（References）

[1] Hassani A, Azapagic A, Shokri N, et al. Predicting long-term dynamics of soil salinity and sodicity on a global scale[J]. Proceedings of the National Academy of Sciences, 2020, 117(52): 33017–33027.

[2] Schoups G, Hopmans J W, Young C A, et al. Sustainability of irrigated agriculture in the San Joaquin Valley, California[J]. Proceedings of the National Academy of Sciences, 2005, 102(43): 15352–15356.

[3] 赵爽, 丁建丽, 韩礼敬, 等. 新疆典型盐渍土微波介电特性响应分析与建模[J]. 干旱区地理, 2022, 45(5): 1534–1546. [Zhao Shuang, Ding Jianli, Han Lijing, et al. Response analysis and modeling of microwave dielectric properties of typical saline soil in Xinjiang[J]. Arid Land Geography, 2022, 45(5): 1534–1546.]

[4] 赵其国, 周生路, 吴绍华, 等. 中国耕地资源变化及其可持续利用与保护对策[J]. 土壤学报, 2006, 43(4): 662–672. [Zhao Qiguo, Zhou Shenglu, Wu Shaohua, et al. Cultivated land resources and strategies for its sustainable utilization and protection in China[J]. Acta Pedologica Sinica, 2006, 43(4): 662–672.]

[5] Zaman M, Shahid S A, Heng L, et al. Guideline for salinity assessment, mitigation and adaptation using nuclear and related techniques[M]. Cham: Springer Nature, 2018: 1–42.

[6] 丁建丽, 王飞. 干旱区大尺度土壤盐度信息环境建模——以新疆天山南北中低海拔冲积平原为例[J]. 地理学报, 2017, 72(1): 64–78. [Ding Jianli, Wang Fei. Environmental modeling of large-scale soil salinity information in an arid region: A case study of the low and middle altitude alluvial plain north and south of the Tianshan Mountains, Xinjiang[J]. Acta Geographica Sinica, 2017, 72(1): 64–78.]

[7] 杨琳, Fahmy Sherif, Hann Sheldon, 等. 基于土壤–环境关系的更新传统土壤图研究[J]. 土壤学报, 2010, 47(6): 1039–1049. [Yang Lin, Fahmy Sherif, Hann Sheldon, et al. Updating conventional soil maps using knowledge on soil-environment relationships extracted from the maps[J]. Acta Pedologica Sinica, 2010, 47(6): 1039–1049.]

[8] Metternicht, Graciela, Alfred Zinck, et al. Remote sensing of soil salinization: Impact on land management[M]. Boca Raton: CRC Press, 2008: 4–56.

[9] Howari F M, Goodell P C, MiyamotoI S. Spectral properties of salt crusts formed on saline soils[J]. Journal of Environmental Quality, 2002, 31(5): 1453–1461.

[10] Mulder V L, De Bruin S, Schaepman M E, et al. The use of remote sensing in soil and terrain mapping: A review[J]. Geoderma, 2011, 162(1–2): 1–19.

[11] Wang J, Ding J, Abulimiti A, et al. Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared (VIS-NIR) spectroscopy, Ebinur Lake Wetland, northwest China[J]. PeerJ, 2018, 6: e4703, doi: 10.7717/peerj.4703.

[12] Elmetwalli A M H, Tyler A N, Hunter P D, et al. Detecting and distinguishing moisture-and salinity-induced stress in wheat and maize through in situ spectroradiometry measurements[J]. Remote Sensing Letters, 2012, 3(4): 363–372.

[13] Gorji T, Yildirim A, Hamzehpour N, et al. Soil salinity analysis of Urmia Lake Basin using Landsat-8 OLI and Sentinel-2A based spectral indices and electrical conductivity measurements[J]. Ecological Indicators, 2020, 112: 106173, doi: 10.1016/j.ecolind.2020.106173.

[14] Fernandez-Buces N, Siebe C, Cram S, et al. Mapping soil salinity using a combined spectral response index for bare soil and vegetation: A case study in the former lake Texcoco, Mexico[J]. Journal of Arid Environments, 2006, 65(4): 644–667.

[15] Ding J, Yu D. Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan-Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments [J]. Geoderma, 2014, 235: 316–322.

[16] Farifteh J, Van der Meer F, Atzberger C, et al. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN)[J]. Remote Sensing of Environment, 2007, 110(1): 59–78.

[17] Grimm R, Behrens T, Märker M, et al. Soil organic carbon concentrations and stocks on Barro Colorado Island: Digital soil mapping using random forests analysis[J]. Geoderma, 2008, 146(1–2): 102–113.

[18] Ma G, Ding J, Han L, et al. Digital mapping of soil salinization based on Sentinel-1 and Sentinel-2 data combined with machine learning algorithms[J]. Regional Sustainability, 2021, 2(2): 177–188.

[19] Allbed A, Kumar L, Aldakheel Y Y, et al. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: Applications in a date palm dominated region[J]. Geoderma, 2014, 230: 1–8.

[20] Wang F, Yang S, Yang W, et al. Comparison of machine learning algorithms for soil salinity predictions in three dryland oases locat-

chinaXiv:202309.00163v1

ChinaXiv合作期刊

chinaXiv:202309.00163v1

ed in Xinjiang Uygur Autonomous Region (XJUAR) of China[J]. European Journal of Remote Sensing, 2019, 52(1): 256–276.

[21] Zhou T, Lu H L, Wang W W, et al. GA-SVM based feature selection and parameter optimization in hospitalization expense modeling[J]. Applied Soft Computing, 2019, 75: 323–332.

[22] Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data[J]. Journal of Bioinformatics and Computational Biology, 2005, 3(2): 185–205.

[23] Peng J, Biswas A, Jiang Q, et al. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China [J]. Geoderma, 2019, 337: 1309–1319.

[24] 马慧琴, 黄文江, 景元书, 等. 基于 AdaBoost模型和mRMR算法的小麦白粉病遥感监测[J]. 农业工程学报, 2017, 33(5): 162–169. [Ma Huiqin, Huang Wenjiang, Jing Yuanshu, et al. Remote sensing monitoring of wheat powdery mildew based on AdaBoost model combining mRMR algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33(5): 162–169.]

[25] Schmid T, Koch M, Gumuzzio J, et al. Applications of hyperspectral imagery to soil salinity mapping[J]. Remote Sensing of Soil Salinization, 2008: 113–139.

[26] Dwivedi R S, Kothapalli R V, Singh A N, et al. Generation of farm-level information on salt-affected soils using IKONOS-II multispectral data[J]. Remote Sensing of Soil Salinization, 2008: 73–90.

[27] Bouaziz M, Matschullat J, Gloaguen R, et al. Improved remote sensing detection of soil salinity from a semi-arid climate in Northeast Brazil[J]. Comptes Rendus Geoscience, 2011, 343(11–12): 795–803.

[28] Eldeiry A A, Garcia L A. Detecting soil salinity in alfalfa fields using spatial modeling and remote sensing[J]. Soil Science Society of America Journal, 2008, 72(1): 201–211.

[29] Frazier A E, Hemingway B L. A technical review of planet smallsat data: Practical considerations for processing and using planetscope imagery[J]. Remote Sensing, 2021, 13(19): 3930, doi: 10.3390/rs13193930.

[30] Roy D P, Wulder M A, Loveland T R, et al. Landsat-8: Science and product vision for terrestrial global change research[J]. Remote Sensing of Environment, 2014, 145: 154–172.

[31] Drusch M, Del Bello U, Carlier S, et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services[J]. Remote Sensing of Environment, 2012, 120: 25–36.

[32] 杨劲松, 姚荣江, 王相平, 等. 中国盐渍土研究: 历程, 现状与展望[J]. 土壤学报, 2022, 59(1): 10–27. [Yang Jingsong, Yao Rongjiang, Wang Xiangping, et al. Research on salt-affected soils in China: History, status quo and prospect[J]. Acta Pedologica Sinica, 2022, 59(1): 10–27.]

[33] Hardie M, Doyle R. Measuring soil salinity[J]. Plant Salt Tolerance, 2012: 415–425.

[34] Aldeghi A, Carn S, Escobar-Wolf R, et al. Volcano monitoring from space using high-cadence planet CubeSat images applied to Fuego volcano, Guatemala[J]. Remote Sensing, 2019, 11(18): 2151, doi: 10.3390/rs11182151.

[35] Xie B, Ding J, Ge X, et al. Estimation of soil organic carbon content in the Ebinur Lake wetland, Xinjiang, China, based on multisource remote sensing data and ensemble learning algorithms[J]. Sensors, 2022, 22(7): 2685, doi: 10.3390/s22072685.

[36] Gorji T, Sertel E, Tanik A, et al. Monitoring soil salinity via remote sensing technology under data scarce conditions: A case study from Turkey[J]. Ecological Indicators, 2017, 74: 384–391.

[37] Huete A R. A soil-adjusted vegetation index (SAVI)[J]. Remote Sensing of Environment, 1988, 25(3): 295–309.

[38] Major D J, Baret F, Guyot G, et al. A ratio vegetation index adjusted for soil brightness[J]. International Journal of Remote Sensing, 1990, 11(5): 727–740.

[39] Liu H Q, Huete A. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise[J]. IEEE Transactions on Geoscience and Remote Sensing, 1995, 33(2): 457–465.

[40] Deering D W, Rouse J W, Haas R H, et al. Measuring forage production of grazing units from Landsat MSS data[J]. Remote Sensing of Environment, 1975(11): 1169–1178.

[41] Khan N M, Rastoskuev V V, Sato Y, et al. Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators[J]. Agricultural Water Management, 2005, 77(1–3): 96–109.

[42] Bannari A, Guedon A M, El-Harti A, et al. Characterization of slightly and moderately saline and sodic soils in irrigated agricultural land using simulated data of advanced land imaging (EO-1) sensor[J]. Communications in Soil Science and Plant Analysis, 2008, 39(19–20): 2795–2811.

[43] Abbas A, Khan S. Using remote sensing techniques for appraisal of irrigated soil salinity[C]//International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, 2007: 2632–2638.

[44] Jing X, Zou Q, Yan J, et al. Remote sensing monitoring of winter wheat stripe rust based on mRMR-XGBoost algorithm[J]. Remote Sensing, 2022, 14(3): 756, doi: 10.3390/rs14030756.

[45] Breiman L. Bagging predictors[J]. Machine Learning, 1996, 24(2): 123–140.

[46] Scudiero E, Skaggs T H, Corwin D L. Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA[J]. Geoderma Regional, 2014, 2: 82–90.

[47] Spies B, Woodgate P. Salinity mapping methods in the Australian context[M]. Canberra: Department of the Environment and Heritage, 2005: 18–45.

[48] 丁建丽, 姚远, 王飞, 等. 干旱区土壤盐渍化特征空间建模[J]. 生态学报, 2014, 34(16): 4620–4631. [Ding Jianli, Yao Yuan, Wang Fei, et al. Detecting soil salinization in arid regions using spectral feature space derived from remote sensing data[J]. Acta Ecologica Sinica, 2014, 34(16): 4620–4631.]

ChinaXiv合作期刊

[49] 曹雷, 丁建丽, 玉米提·哈力克, 等. 基于国产高分一号卫星数据的区域土壤盐渍化信息提取与建模[J]. 土壤学报, 2016, 53(6): 1399-1409. [Cao Lei, Ding Jianli, Halik Umut, et al. Extraction and modeling of regional soil salinization based on data from GF-1 satellite[J]. Acta Pedologica Sinica, 2016, 53(6): 1399-1409.]
[50] Rao B R M, Sharma R C, Ravi Sankar T, et al. Spectral behaviour of salt-affected soils[J]. International Journal of Remote Sensing, 1995, 16(12): 2125-2136.
[51] 冯娟, 丁建丽, 杨爱霞, 等. 干旱区土壤盐渍化信息遥感建模[J]. 干旱地区农业研究, 2018, 36(1): 266-273. [Feng Juan, Ding Jianli, Yang Aixia, et al. Remote sensing modeling of soil salinization information in arid areas[J]. Agricultural Research in Arid Areas, 2018, 36(1): 266-273.]
[52] Metternicht G I, Zinck J A. Remote sensing of soil salinity: Potentials and constraints[J]. Remote Sensing of Environment, 2003, 85(1): 1-20.
[53] 丁建丽, 伍漫春, 刘海霞, 等. 基于综合高光谱指数的区域土壤盐渍化监测研究[J]. 光谱学与光谱分析, 2012, 32(7): 1918-1922. [Ding Jianli, Wu Manchun, Liu Haixia, et al. Study on the soil salinization monitoring based on synthetical hyperspectral index[J]. Spectroscopy and Spectral Analysis, 2012, 32(7): 1918-1922.]
[54] 马国林, 丁建丽, 韩礼敬, 等. 基于变量优选与机器学习的干旱区湿地土壤盐渍化数字制图[J]. 农业工程学报, 2020, 36(19): 124-131. [Ma Guolin, Ding Jianli, Han Lijing, et al. Digital mapping of soil salinization in arid area wetland based on variable optimized selection and machine learning[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(19): 124-131.]

chinaXiv:202309.00163v1

Digital mapping of soil salinization in a typical oasis based on PlanetScope images

LI Ke[1,2], DING Jianli[1,2], HAN Lijing[1,2], GE Xiangyu[1,2], GU Yongsheng[1,2], ZHOU Qian[1,2], LYU Yangxia[3]

(1. Xinjiang Common University Key Lab of Smart City and Environment Stimulation, College of Geography and Remote Sensing Sciences, Xinjiang University, Urumqi 830046, Xinjiang, China; 2. Key Laboratory of Oasis Ecology under Ministry of Education, Xinjiang University, Urumqi 830046, Xinjiang, China; 3. Groundwater and Salt Monitoring Station, Ogan River Basin Management Office, Aksu 842000, Xinjiang, China)

Abstract: High-resolution soil salinity maps are urgently needed in arid and semi-arid regions to visualize the subtle spatial variations in salinity distribution. These maps are crucial for guiding the development of land resource management policies and water resource management policies in salt-affected and potentially salt-affected areas, aiming to prevent further soil degradation and ensure sustainable agricultural economic development and food security. Based on PlanetScope imagery, vegetation spectral indices and soil salinity indices were extracted, resulting in a total of 21 variables. These variables were used as input for the Bagging algorithm to construct a soil salinity prediction model, referred to as Model-Ⅰ. The max-relevance and min-redundancy (mRMR) method was employed to select relevant feature variables, which were then inputted into the Bagging algorithm to build a soil salinity prediction model, referred to as Model-Ⅱ. Field sampling data were used to assist in model building and validation. Model-Ⅰ and Model-Ⅱ were evaluated using model evaluation metrics. The results indicate that the prediction performance of Model-Ⅱ is better than that of Moedl-Ⅰ (mean R^2=0.66, mean RMSE=18.00 $dS \cdot m^{-1}$, mean PRIQ=3.21 for the validation set), and that mRMR effectively reduces the multidimensional feature redundancy. PlanetScope images combined with the mRMR method successfully mapped high-resolution soil salinity, which provided more detailed information on the spatial distribution of soil salinity, and the results of the study promoted the use of PlanetScope data to monitor soil salinity information.

Key words: PlanetScope; mRMR; Bagging; soil salinization; digital mapping techniques